Master Data Science: Latent variable modelling - assignement 4

contact: jean-marc.freyermuth@univ-amu.Fr

Academic year 2020-2021

Exercice 1: Bayesian hierarchical model

The table hereafter describes the multiple failures of pumps in a nuclear plant. We consider that the failures of the i-th pump follow a Poisson process with parameter λ_i. For an observed time t_i; the number of failures y_i is thus a Poisson $Poi(\lambda_i t_i)$ random variable. The prior distributions are:

$$\begin{aligned}&\lambda_i|\beta \sim \mathcal{G}(\alpha, \beta),\ 1 \leq i \leq 10\\ &\beta \sim \mathcal{G}(\gamma, \delta)\\ &\perp_{1 \leq i \leq n} \lambda_i|\beta.\end{aligned}$$

with $\alpha = 1.8$, $\gamma = 0.01$ $\delta = 1$.

Pump / number of failures / time (in thousand of hours)

```r
nb_failures = c(5,1,5,14,3,19,1,1,4,22)
time = c(94,16,63,126,5,31,1,1,2,10)
pump =1:10
data_pumps = data.frame(pump,nb_failures,time)
data_pumps
```

```
##    pump nb_failures time
## 1     1           5   94
## 2     2           1   16
## 3     3           5   63
## 4     4          14  126
## 5     5           3    5
## 6     6          19   31
## 7     7           1    1
## 8     8           1    1
## 9     9           4    2
## 10   10          22   10
```

- determine the joint posterior distribution and the full conditional distributions
- provide 95% credibility intervals for the λ_i's and for β.
- consider that this set of 10 pumps are involved in the cooling system of a nuclear power plant. The system is working if at least one pump is working. Compute the probability that the system will work well more than 10000 hours (Pumps will not be repaired).
- Answer the same question but you now consider the following prior

$$\begin{aligned}&\lambda_i|\alpha,\beta \sim \mathcal{G}(\alpha,\beta),\ 1\leq i\leq 10\\&\beta\sim\mathcal{G}(\gamma,\delta)\\&\alpha\sim\mathcal{U}(0,5)\end{aligned}$$

with $\gamma=0.01$ $\delta=1$.

Exercice 2: Change point detection in count time series

The following data set gives the number of coalmine disasters in UK from 1851 to 1962 denoted hereafter as $\{y_t,\ 1851\leq t\leq 1962\}$.

Source: Carlin, Gelfand, and Smith (1992) Hierarchical Bayesian Analysis of Changepoint Problems Applied Statistics volume 41, pages 389-405.

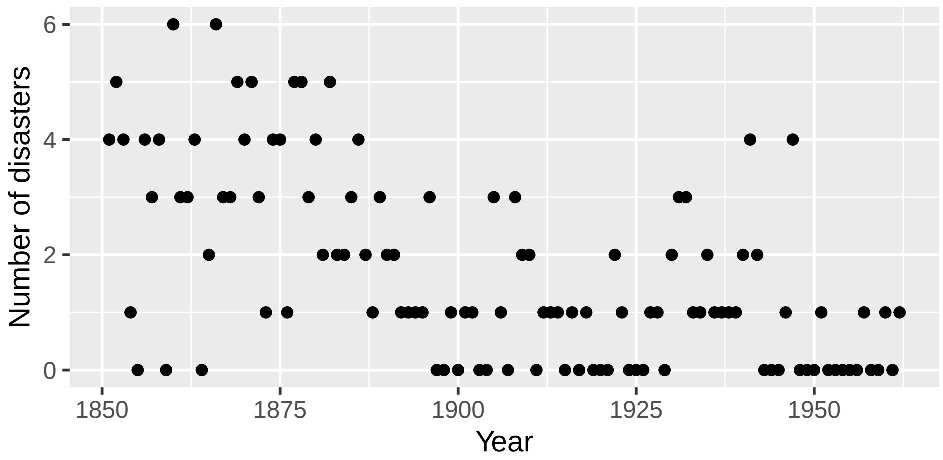
```
##   Year Count
## 1 1851     4
## 2 1852     5
## 3 1853     4
## 4 1854     1
## 5 1855     0
## 6 1856     4
```

Looking at these data you may think that there is a sudden change in the rate of occurrence of disasters at a year *lambda*. This can be due to changes in policy (more mine inspections), in safety rules…

```
library(ggplot2, quietly = TRUE)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.3
```

```
ggplot(CoalDisast)+aes(x=Year,y=Count)+ xlab("Year") + ylab("Number of disasters") + geom_point()
```



Hence you consider the following change-point model for you data:

$$y_t|\gamma,\delta,\lambda=\left\{\begin{array}{l}Poi(\gamma),\ \text{si } t\leq\lambda,\\ Poi(\delta),\ \text{si } t>\lambda,\end{array}\right.$$

where $Poi(\tau)$ is Poisson distribution with mean τ. Let,

$$\begin{aligned} y_t|\tau &\sim Poi(\tau) \\ p(y_t|\tau) &= \frac{\tau^{y_t}}{y_t!}e^{-\tau}. \end{aligned}$$

You suggest the following prior distributions

$$\begin{aligned} &\gamma \sim \mathcal{G}(a_1, a_2), \\ &\delta \sim \mathcal{G}(b_1, b_2), \\ &\lambda \sim \mathcal{U}\{1, 2, \dots, (T-1)\} \end{aligned}$$

- Explain your choice for using Gamma prior on the parameters γ et δ.
- Write down the likelihood of the model.
- Compute the posterior distribution.
- Based on these data, give the most probable value of λ, i.e., the date of a sudden regime change.

Help 1: Let g be a discrete distribution on $0, 1, \dots, K$. If we know the form of g up to a multiplicative constant then we can generate observations from g from its renormalized form.

Help 2: to sample randomly (eventually according to a vector of weight) among a set of values, you can use the function *sample*.

- Give the probability that λ is greater than $\delta + 2$ in between two period of time.